

(11)Publication number : 2001-188777
(43)Date of publication of application : 10.07.2001

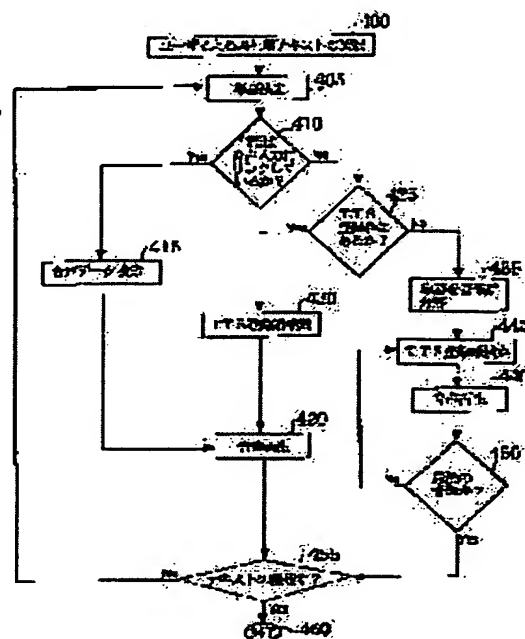
606F 17/22
606F 3/16
606F 17/28
610L 13/00
610L 15/06
610L 15/00

(71)Applicant : MICROSOFT CORP
(72)Inventor : REYNAR JEFFREY C
RUCKER ERICK
KIM PAUL KYONG HWAN

Priority number : 1999 428259 Priority date : 27.10.1999 Priority country : US

(57)Abstract:

SOLUTION: An input is accepted from various sources, and the inputs are integrated into a text, and the part selected by a user in the text is read. The text is displayed as a document in a word processor, and spoken data are preserved in a short term or long term storage device as retrieval voice data. The text can be freely edited, changed, and processed by the user. The user can select one part of the text, and request the reproduction of a voice. Then, each word of the selected text is checked, and when the preserved voice data are related with the word, the voice data are extracted and reproduced. When any related voice data are not present, registered words integrated into a text voice are retrieved, and the retrieved words are reproduced instead.



(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-188777

(P2001-188777A)

(43) 公開日 平成13年7月10日 (2001.7.10)

(51) Int.Cl.⁷

識別記号

F I

テマコード* (参考)

G 0 6 F 17/22

5 0 3

G 0 6 F 17/22

5 0 3

3/16

3 3 0

3/16

3 3 0 G

17/28

17/28

V

G 1 0 L 13/00

G 1 0 L 3/00

E

15/06

5 2 1 C

審査請求 未請求 請求項の数31 O L (全 13 頁) 最終頁に続く

(21) 出願番号 特願2000-329016(P2000-329016)

(22) 出願日 平成12年10月27日 (2000. 10. 27)

(31) 優先権主張番号 0 9 / 4 2 8 2 5 9

(32) 優先日 平成11年10月27日 (1999. 10. 27)

(33) 優先権主張国 米国 (U S)

(71) 出願人 595124929

マイクロソフト コーポレイション

MICROSOFT CORPORATI
ON

アメリカ合衆国 98052-6399 ワシント
ン州 レドモンド ワン マイクロソフト
ウェイ (番地なし)

(72) 発明者 ジェフリー スイー. レイナー

アメリカ合衆国 98072 ワシントン州
ウッドインヴィル エヌイー ワンハンド
レッドサーティエイス ウェイ 23322

(74) 代理人 100083806

弁理士 三好 秀和 (外1名)

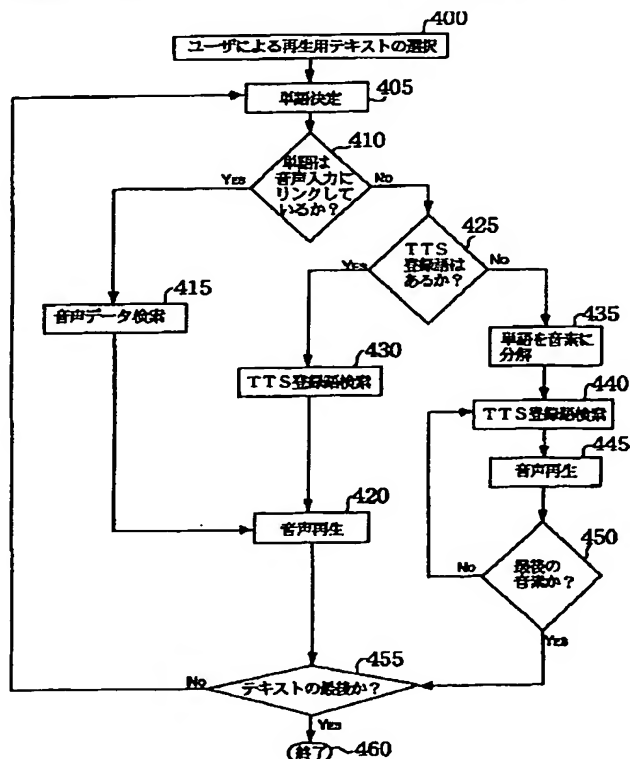
最終頁に続く

(54) 【発明の名称】 音声テキストに関連付ける方法、音声テキストに関連付けるコンピュータ、コンピュータで文書を生成し読み上げる方法、文書を生成し読み上げるコンピュータ、コンピュータでテキスト

(57) 【要約】

【課題】 文書を選択部分に対応する音声を直感的な形式で確実に再生し、音声データが欠落した部分を穴埋めして音声再生する方法を提供する。

【解決手段】 多様なソースから入力を受け入れ、それら入力をテキスト化し、そのテキストの中でユーザが選択した箇所を読み上げる。ワードプロセッサ内の文書としてテキストを表示し、口述データを検索可能音声データとして短期または長期記憶装置に保存する。このテキストは、ユーザによって自由に編集、変更、処理できる。ユーザは前記テキストの一部を選択し音声再生を要求できる。選択されたテキストの各語を調べ、その語に対して保存音声データが関連付けられていればその音声データを取り出し再生する。関連付けられた音声データが無ければ、テキスト音声化登録語を検索し、検索した語を代わりに再生する。



【特許請求の範囲】

【請求項1】 語を口述するステップと、
文書中の第1テキストとして前記口述語を文字化するステップと、
前記口述を記憶媒体に保存するステップと、
前記口述の各語と前記テキストの要素とを関連付けるステップと、
前記文書に第2テキストを挿入するステップと、
前記第2テキストにテキスト音声化登録語を関連付けるステップとを有することを特徴とする音声テキストに関連付ける方法。

【請求項2】 前記文書内における前記第1テキストと第2テキストの配置順序に従って前記口述とテキスト音声化登録語とを再生するステップをさらに有することを特徴とする請求項1に記載の方法。

【請求項3】 前記第2テキストを第1口述語と第2口述語との間に挿入することを特徴とする請求項2に記載の方法。

【請求項4】 主第2テキストが前記口述語に先行し、副第2テキストが前記口述語に続くことを特徴とする請求項2に記載の方法。

【請求項5】 主第1テキストと副第1テキストとが、主第2テキストと副第2テキストとに交替することを特徴とする請求項2に記載の方法。

【請求項6】 語を口述する手段と、
文書中の第1テキストとして前記口述語を文字化する手段と、
前記口述を記憶媒体に保存する手段と、
前記口述の各語と前記テキストの要素とを関連付ける手段と、
前記文書に第2テキストを挿入する手段と、
前記第2テキストにテキスト音声化登録語を関連付ける手段とを有することを特徴とする音声をテキストに関連付けるコンピュータ。

【請求項7】 前記文書内における前記第1テキストと第2テキストの配置順序に従って前記口述とテキスト音声化登録語とを再生する手段をさらに有することを特徴とする請求項6に記載のコンピュータ。

【請求項8】 入力装置に少なくとも1つの語を発声入力するステップと、
前記語を文字化して文書に第1テキスト登録語として転写するステップと、
前記語を記憶媒体に保存するステップと、
前記語を前記第1テキスト登録語に関連付けるステップと、
前記文書内に第2テキスト登録語をタイピング入力するステップと、
前記第2テキスト登録語に対応するテキスト音声化登録語を前記第2テキスト登録語に関連付けるステップと、
前記文書内における前記第1および第2テキスト登録語

の順序に従って前記語とテキスト音声化登録語とを再生するステップとを有することを特徴とするコンピュータで文書を生成し読み上げる方法。

【請求項9】 前記語の音声特長を分析し、前記音声特長に合わせて前記テキスト音声化登録語の再生を調整するステップをさらに有することを特徴とする請求項8に記載の方法。

【請求項10】 前記テキスト音声化登録語の韻律要素が調整されることを特徴とする請求項9に記載の方法。

【請求項11】 ディスプレイ画面に前記文書を表示するステップと、前記テキスト音声化登録語に対応する前記語に影を付けるステップとをさらに有することを特徴とする請求項8に記載の方法。

【請求項12】 前記第1テキスト登録語が停止し前記第2テキスト登録語が開始する信号が非英数字文字の表示によって送られることを特徴とする請求項8に記載の方法。

【請求項13】 前記第1および第2テキスト登録語が絵文字を有することを特徴とする請求項8に記載の方法。

【請求項14】 前記絵文字が漢字であることを特徴とする請求項13に記載の方法。

【請求項15】 入力装置に少なくとも1つの語を発声入力する手段と、
前記語を文字化して文書に第1テキスト登録語として転写する手段と、
前記語を記憶媒体に保存する手段と、
前記語を前記第1テキスト登録語に関連付ける手段と、
前記文書内に第2テキスト登録語をタイピング入力する手段と、
前記第2テキスト登録語に対応するテキスト音声化登録語を前記第2テキスト登録語に関連付ける手段と、
前記文書内における前記第1および第2テキスト登録語の順序に従って前記語とテキスト音声化登録語とを再生する手段とを有することを特徴とする文書を生成し読み上げるコンピュータ。

【請求項16】 前記韻律要素がピッチと速度と音量であることを特徴とする請求項10に記載の方法。

【請求項17】 前記第1および第2テキスト登録語が、周囲テキストに応じて形状を変化させることを特徴とする請求項8に記載の方法。

【請求項18】 前記第1および第2テキスト登録語が右から左へ読まれることを特徴とする請求項8に記載の方法。

【請求項19】 少なくとも1つの音素からなる少なくとも1つの語で、構成したテキストセットを選択するステップと、
ユーザの口述音声入力第1語に対応するか否かを決定するステップと、
ユーザの口述音声入力第1語に対応すれば、音声出力

装置を介して前記ユーザの口述音声入力を再生するステップと、

ユーザの口述音声入力が第1語に対応しなければ、複数のテキスト音声化登録語の1つが前記第1語に対応するか否かを決定するステップと、

1つのテキスト音声化登録語が前記第1語に対応すれば、音声出力装置を介して前記テキスト音声化登録語を再生するステップと、

1つのテキスト音声化登録語が前記第1語に対応しなければ、前記複数のテキスト音声化登録語のどれが前記音素に対応するかを決定するステップと、

前記複数のテキスト音声化登録語のどれが第1音素に対応するか決定に応じて、音声出力装置を介して前記対応するテキスト音声化登録語を再生するステップとを有することを特徴とするコンピュータでテキスト文書の音声再生を行う方法。

【請求項20】 前記テキストセットが複数の語からなり、前記第1語がユーザの口述音声入力に対応し、第2語がテキスト音声化登録語に対応することを特徴とする請求項19に記載の方法。

【請求項21】 前記テキストセット内における前記第1語および第2語の順序に従って前記ユーザの口述音声入力とテキスト音声化登録語とを再生するステップをさらに有することを特徴とする請求項19に記載の方法。

【請求項22】 対応するユーザの口述音声入力がない複数の語を決定するステップと、前記複数の語をテキスト音声化モジュールへ渡すステップと、前記複数の語の各々についてテキスト音声化登録語を検索するステップとをさらに有することを特徴とする請求項19に記載の方法。

【請求項23】 少なくとも1つの音素からなる少なくとも1つの語で、構成したテキストセットを選択する手段と、

ユーザの口述音声入力が第1語に対応するか否かを決定する手段と、

ユーザの口述音声入力が第1語に対応すれば、音声出力装置を介して前記ユーザの口述音声入力を再生する手段と、

ユーザの口述音声入力が第1語に対応しなければ、複数のテキスト音声化登録語の1つが前記第1語に対応するか否かを決定する手段と、

1つのテキスト音声化登録語が前記第1語に対応すれば、音声出力装置を介して前記テキスト音声化登録語を再生する手段と、

1つのテキスト音声化登録語が前記第1語に対応しなければ、前記複数のテキスト音声化登録語のどれが前記音素に対応するかを決定する手段と、前記複数のテキスト音声化登録語のどれが第1音素に

対応するか決定に応じて、音声出力装置を介して前記対応するテキスト音声化登録語を再生する手段とを有することを特徴とするテキスト文書の音声再生を行うコンピュータ。

【請求項24】 音声認識処理を用いて口述語を文書内の第1テキストセットに変換するステップと、後からの音声再生のため、前記口述語を保存すると共に前記第1テキストに関連付けるステップと、非音声命令を前記文書内の第2テキストセットに変換するステップと、前記文書内における前記第1および第2テキストセットの順序に従って前記第1および第2テキストセットに対応する音声を再生し、前記音声の前記第1テキストセットに対応する部分は前記保存口述語の再生によって提供し、前記音声の前記第2テキストセットに対応する部分はテキスト音声化処理によって提供するステップとを有することを特徴とする文書内のテキストを編集し評価する方法。

【請求項25】 キーボードを用いて前記非音声命令を供給することを特徴とする請求項24に記載の方法。

【請求項26】 マウスを用いて前記非音声命令を供給することを特徴とする請求項24に記載の方法。

【請求項27】 口述語からテキストへの前記変換を音声認識処理を用いて行うことを特徴とする請求項24に記載の方法。

【請求項28】 前記第2テキストセットの音声再生をテキスト音声化処理を用いて行うことを特徴とする請求項24に記載の方法。

【請求項29】 手の動作によって前記非音声命令を供給することを特徴とする請求項24に記載の方法。

【請求項30】 筆記タブレットを用いて前記非音声命令を供給することを特徴とする請求項24に記載の方法。

【請求項31】 現在の再生に対応する視覚的手掛かりをディスプレイ画面に表示することを特徴とする請求項24に記載の方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は広く音声認識分野に関する。さらに詳しくは、コンピュータ表示用のマルチソース入力再生ユーティリティに関する。

【0002】

【従来の技術】パーソナルコンピュータの出現以来、コンピュータとのヒューマンインタフェースは基本的にキーボードであった。例えばコンピュータに情報を入力する場合、ユーザはそのコンピュータに付属するキーボードで情報をタイプする。キーボードを補うため、他の入力機器も使われてきた。例えばマウス、タッチパネルディスプレイ、統合ポインタ機器、スキャナである。ユーザはこれら入力機器を使うことにより、データやコマンドをコンピュータに入力する時間を短縮できた。

【0003】コンピュータベースの音声認識や音声認識システムも、パーソナルコンピュータへのデータ入力およびコマンド入力に使われてきた。音声認識および音声認識システムは、人の音声コンピュータが理解できる形式に変換する。コンピュータが音声認識システムを備えれば、入力装置にデータを話すだけでデータ入力が行われる。ユーザの会話速度は、従来のデータ入力より一般に速い。従ってパーソナルコンピュータに音声認識および音声認識システムを組み込むことの利点は、会話によるデータ入力の速度にある。

【0004】音声認識および音声認識システムを備えたパーソナルコンピュータは、ユーザの効率を高めるため、職場での利用が進んでいる。音声認識および音声認識システムは、様々な産業分野で様々な形で適用されている。例えばドラゴン社、アイビーエム社、ランアウトアンドハウスパイ社は、音声認識および音声認識技術を用いたコンピュータソフトウェアプログラムを開発した。音声認識プログラムに対してユーザが文書を音読したり口述すると、その音声認識プログラムは、ワードプロセッサ等のパーソナルコンピュータで動作するアプリケーションにユーザの話した語を直接入力する。

【0005】コンピュータベースの音声認識および音声認識プログラムは、一般に人の音声を一連のデジタル周波数に変換する。これら周波数は、あらかじめ記憶した音素と呼ばれる語や音声要素と照合される。

【0006】音素は、音声の中の各音を区別するための最小単位である。各音素は、1つ以上の対応する異音を有する。異音は、音素の音声表現である。1つの音素は、多くの異音を有することがある。この場合、各異音は、語中の音素の位置や、同一文字セットの言語における異形発音により、わずかに異なった音を有する。例えば音素「b」は、「boy (ボーイ)」と「beyond (ビヨンド)」とで発音が異なる。この時の各発音が音素「b」の異音である。

【0007】前記ユーティリティは、これら音素を処理し、各音素に最適なテキスト表現に基づいて、音素をテキストに変換する。その方法は当業者に良く知られている。音素から変換したテキストは、ワードプロセッサ、スプレッドシート、データベース、ウェブブラウザ等、音声入力を受け取り、それを表示テキストやプログラムコマンドに変換できるプログラムによって表示する。前記マルチソース入力再生ユーティリティは、音声データを記憶することもできる。音声データは様々な形式で様々な記憶媒体に記憶できる。例えば揮発性RAM、長期保存用磁気記憶装置、CD-ROM等の光媒体に記憶できる。音声データは、さらに圧縮して記憶容量を縮小しても良い。前記ユーティリティは、記憶した音声データをその音声データから生成したテキストに関連付け、後からの再生に利用できる。コンピュータが一連の周波数の一致を正しく決定すれば、その部分の音声のコンピュ

ータ認識は完了する。周波数一致処理は、必要な情報がコンピュータに集まるまで続ける。コンピュータは、音声記憶装置に格納し、話された単語に応答し、その音声を文書のテキストに変換する。このテキストは、ワードプロセッサプログラムで扱える形式、あるいはアプリケーションプログラムの実行コマンドである。

【0008】最終的に市場に出回ることが期待されているのは、自然会話入力システムである。このシステムは、ユーザが特別の方法で喋らなくとも、コンピュータへの音声入力を可能とし、コンピュータへのユーザコマンドと入力情報とを区別できる。

【0009】現在の音声認識および音声認識システムは、この進歩的技術に欠けているため、信頼性が足りない。ハードウェアおよびソフトウェアの改良にもかかわらず、最良の音声認識および音声認識システムでさえ、97～99%の信頼性しか確保できない。内的および外的要因が音声認識および音声認識システムの信頼性に影響する。認識技術自体の要因は、話者の言語における単語や音素の数が有限なことであり、音声認識ソフトの話者入力照合用語彙が有限なことである。地域的アクセント、外部騒音、マイクロフォン品質等の環境要因は入力品質を低下させるので、ユーザ音読語の周波数に影響を与え、語照合や音素照合にエラーを生ずることがある。

【0010】このため、音声認識ソフトウェアがテキスト化した口述文書は、認識エラーを含むことが多い。タイピングエラーの場合、文字の移動などの単純ミスは、特定したり訂正することが容易である。ところが認識エラーは深刻な場合が多い。認識エラーは、文字の置き換えや移動ではなく、同じ発音の単語の置き換えであることが多い。例えば古くからある音声認識エラーは、「音声認識 (レコグナイズスピーチ、recognize speech)」を「素敵な海辺を破壊 (レッカナイズスピーチ、wreck a nice beach)」とテキスト化してしまうことがある。これらは発音が似ているものの、意味が全く異なる。この種の認識エラーを含む文書を校正する場合、即座に本来の語句を思い出すことが難しく、余計な混乱を招く。

【0011】従来このような混乱を最小にするため、音読しながら文書を校正してきた。このやり方は、本来の語句を特定する助けになる。音読すれば発音の類似性が明らかになるからだ。しかし文書の口述と編集との間に長い時間が空くと、ユーザは本来の語句が何であったかを忘れがちである。

【0012】従来の音声認識製品は、この問題を解決するため、口述を音声データとして保存し、テキスト化した各語にその保存音声データを関係付ける。ユーザは単語または一連のテキストを選択し、その選択部分に対応する音声の再生を要求する。

【0013】これはユーザが本来のテキストを認識するための助けになるが、口述時と音声再生要求時との間に

ユーザが文書を編集した場合、深刻な問題が起きる。ユーザは、混在した入力ソースの文書の一部を再生要求し、その提示を受ける。

【0014】例えばユーザが「私はコンピュータが音声を認識できたら良いと思う (I wish my computer could recognize speech)」と口述したとする。それを音声認識システムが「私はコンピュータが素敵な海辺を破壊できたら良いと思う (I wish my computer could wreck a nice beach)」とテキスト化したとする。次にユーザが「私 (I)」と「良いと思う (wish)」との間に「本当に (really)」とタイプしたとすれば、この文書は混在した入力ソースを有する。ユーザが画面上で「私はコンピュータが素敵な海辺を破壊できたら良いと本当に思う (I really wish my computer could wreck a nice beach)」を選択し再生を要求すると、「本当に (really)」は口述ではなくタイプされたために、それに関連付けられた音声データは無い。

【0015】この状況において、従来の音声認識システムは、再生オプションを無効にする。あるいは、選択テキストに対する音声データが存在しないため再生不可能である、とのエラーメッセージをユーザに返す。これによりユーザは、文書のどの部分をタイプしどの部分を口述したかを思い出し、それから文書を再選択しなければならない。この方法はユーザをいらいらさせる。なぜならユーザは、曖昧な記憶から口述を思い出し、音声再生を行わねばならないからである。

【0016】

【発明が解決しようとする課題】このように本技術分野においては、文書を選択部分に対応する音声を直感的な形式で確実に再生する方法およびシステムが求められている。また、音声データが欠落した部分を有する文書の音声再生において、その欠落部分を穴埋めする方法およびシステムが求められている。

【0017】本発明は上記事情に鑑みてなされたものであり、その目的とするところは、音声をテキストに関連付ける方法、コンピュータで文書を生成し読み上げる方法、コンピュータでテキスト文書の音声再生を行う方法、及び、文書内のテキストを編集し評価する方法を提供することにある。

【0018】本発明の目的は、音声をテキストに関連付けるコンピュータ、文書を生成し読み上げるコンピュータ、及び、テキスト文書の音声再生を行うコンピュータを提供することにある。

【0019】

【課題を解決するための手段】本発明は広く、パーソナルコンピュータ用マルチソース入力再生ユーティリティである。このユーティリティは、複数の入力ソースから

入力を受け入れ、それら入力をテキストに変換し、そのテキストをディスプレイ画面に表示する。ユーザがテキストを口述すると、ユーティリティはその音声入力を音声データとして記憶媒体やシステムメモリに保存する。音声入力から生成したテキストは、この保存音声データに関連付ける。筆記タブレットから文字化したテキスト、あるいはキーボードでタイプしたテキストは、音声データとの関連付けが無い。ユーザはキーボード、マウス等の入力機器を用いて前記テキストを必要に応じて編集する。編集による変更は、一般にテキストへの直接タイピングによって行うので、その変更に関連付けた音声データは保存されていない。

【0020】前記マルチソース入力再生ユーティリティは、ユーザが選択したテキストを音声化する。選択したテキストの全てが保存音声データに関連付けられていれば、その音声データを再生する。選択テキストの一部が関連付け音声データを持っていなければ、前記ユーティリティはテキスト音声化 (TTS) 登録語を検索し、検索した登録語によって保存音声データの再生を穴埋めする。従ってユーザがマルチソーステキストの一部を選択して再生要求した場合、その部分の音声化は、利用可能な再生用音声データと、必要に応じて導入したテキスト音声化登録語との混在によって実現する。

【0021】本発明は、前記必要性に応えるため、非音声入力を含む複数の入力方法を用いて入力したテキストを音声化するための簡単な方法を提供する。本方法は、マルチソース文書において、関連付けられた音声データを持たない語についても、テキスト音声化登録語を検索することにより、エラーメッセージを出すことなく、その文書をコンピュータに音読させる。さらに本方法は、全選択テキストの連続的再生を実現し、非口述テキスト部をスキップすることがないため、ユーザの混乱を最小にする。

【0022】

【発明の実施の形態】本発明の実施の形態は、マルチソース入力再生ユーティリティである。このユーティリティは、様々な入力機器から様々なフォーマットのユーザ入力を受け取り、そのユーザ入力をテキストに変換する。ユーザは、キーボード (タイプしたデータ用)、マイクロフォン (音声データ用)、スタイラスパッド (手書き用)、タッチパッドまたはマウス (データグループからの要素選択用) 等の様々な入力機器を用いて様々なデータフォーマットを入力できる。

【0023】ユーザ入力から変換した文書テキストは、ディスプレイ装置に表示できる。表示装置は例えば、CRTモニター、タッチスクリーン、LCDディスプレイである。ユーザは表示された文書テキストの一部を選択し、音声再生要求を開始する。

【0024】本マルチソース入力再生ユーティリティは、アクセント、会話速度、背景騒音等の様々な要因に

より、ユーザの会話を誤認することがある。これら要因により、前記ユーティリティは、ユーザが喋った異音に一致しない音素を選択してしまう。すると文字化したテキストは不正確になる。

【0025】テキストが口述から大きく食い違うと、テキスト化した文書を見るだけでは元々何を口述したかを思い出すのは難しい。この場合、ユーザは画面上のテキストの一部を選択し、その選択テキストの元となった音声を読み上げるよう、マルチソース入力再生ユーティリティに要求する。ユーザがテキストの一部を再生用を選択すると、前記ユーティリティはその選択されたテキストをハイライト表示し、ユーザに対して視覚的なフィードバックを行う。このフィードバックは、選択テキストを異なる色で表示したり、点滅させたりする事でも良い。選択テキストを他の部分よりも目立たせれば良い。

【0026】前記ユーティリティは次に、選択テキストに関連付けた音声データを検索し、その音声データをスピーカから再生する。音声データ再生中の文書部分に、その箇所を明示する視覚的手掛かりを表示しても良い。例えば音声再生中の語の上にカーソルやボールを表示し、その表示を再生に従って移動させても良い。再生中の音声に対応する画面上の語を点滅させたり、その語に下線を付けても良い。

【0027】ユーザは、口述文書の編集、新しいテキストの追加、音声認識エラーの訂正を希望することがある。この場合、再び音声モジュールを使うこともできるが、通常は口述によらず、キーボードを用いて前記マルチソース入力再生ユーティリティに変更を直接入力し、文書の編集を行う。タッチパッドや筆記スタイラス等の装置によって文書を訂正することもできる。

【0028】このように、文書はマルチソース入力を含むことができる。文書の一部は音声を使って生成したものであり、他の部分はタイピングや手書きによって生成したものである。

【0029】ユーザはマルチソース入力を含む文書の一部の音声再生を要求できる。するとマルチソース入力再生ユーティリティは、選択部分のうち音声データに関連付けられている語と、そうでない語とを決定する。この決定を行うに当たり、選択部分の最初から最後まで各語を順次調べる。あるいは、音声データを利用可能なテキスト部分にポインタを設定し、対応する語グループの記憶場所を示すことにより、語ごとのチェックを省略することもできる。文書の選択部分内の語に関連付けた音声データが無ければ、その語に対応するテキスト音声化(TTS)登録語が存在するか否かを決定する。対応するTTS登録語があれば、その登録語を取り出し、再生する。対応するTTS登録語が無ければ、その語を音素に分解し、その音素に最適な異音に対応するTTS登録語を検索し、その登録語を再生する。例えば姓「ヘメンウェイ(Hemenway)」は対応するTTS登録語

が無いであろう。この場合、その語を音素に分解し、その音素に最適な異音を選択する。そして対応するTTS登録語を選択し、再生する。この手順はその語の全ての音素を音声化するまで続行する。

【0030】本マルチソース入力再生ユーティリティは、選択テキストを解析し、保存音声データに関連付けられていない全ての語を決定し、これら関連付けの無い語に近似のTTS登録語を1回で決定することもできる。すなわちTTSモジュールへの複数回のアクセスをしない。この方法は、TTSアクセスを最小に保つことによってCPU使用を最小に押さえる。この方法は、音声データに関連付けられている語の全てを取り出し、システムメモリに保持し、その間に各非関連付け語に対応するTTS登録語を獲得する。その後、前記音声データとTTS登録語とを順次システムメモリから再生する。すなわち記憶媒体から音声データとTTS登録語とを逐次読み出すことをしない。

【0031】本ユーティリティは、再生前にTTS登録語を速めたり遅くしたりすることにより、それを保存音声データに合わせることができる。音声データ波形を分析し、音声速度、ピッチ、トーン、音色等の情報を抽出しても良い。その後、TTS登録語の対応する特性を変化させることにより、話者の会話パターンにTTS登録語の音声を近似できる。

【0032】変形した実施の形態として、テキストの口述部分を元通りに音声再生しても良い。本実施の形態は、「私はコンピュータが素敵な海辺を破壊できたら良いと本当に思う(I really wish my computer could wreck a nice beach)」を、「私はコンピュータが音声を認識できたら良いと思う(I wish my computer could recognize speech)」と再生する。文書に対する編集改変やタイプによる追加が多いと、干渉が多すぎて、ユーザは音声再生を追跡できないからである。

【0033】他の実施の形態は、タイプした語に対応する箇所を再生する時、一時停止を挿入する。「私はコンピュータが音声を認識できたら良いと思う(I wish my computer could recognize speech)」を編集して「私(I)」と「良いと思う(wish)」の間に「本当に(really)」を挿入した場合を考える。本実施例は、選択テキストを再生する時、「私は(I)」 「一時停止」 「コンピュータが音声を認識できたら良いと思う(wish my computer could recognize speech)」と読み上げる。この一時停止は、タイプした語「本当に(really)」の存在を示す。一時停止箇所に到達すると音声が消えるので、ユーザはそれを知ることができる。一時停止の期間を変えることにより、口述文書に挿入した文や段落を示唆する

こともできる。

【0034】アプリケーションの例としてワードプロセッサを取り上げ本発明を説明したが、本発明は他の多くのアプリケーションに適用可能である。例えばスプレッドシート、ブラウザ、電子メールプログラム、音楽編曲プログラム、CADプログラム、オペレーティングシステム等、当業者には明白なアプリケーションである。

【0035】本マルチソース入力再生ユーティリティは、以下に説明する様々なプログラムモジュールからなる。当業者には明らかなように、これらモジュールは様々な方法で組合せ可能であると共に、新しいプログラムモジュールを作り出すことによって同様な結果をもたらすことも可能である。例えば入力インタフェースやテキスト音声化モジュールは、オペレーティングシステムの一部、ワードプロセッサの一部、あるいはスタンドアロンでも良い。これについては後述する。また手書き認識プログラムモジュール、音声認識プログラムモジュール、入力システムは、独立したプログラムでも良く、ワードプロセッサまたはオペレーティングシステムの部品でも良く、これら3つをスタンドアロンにしてそれらをパッケージにしたものでも良い。

【0036】(動作環境例) 本発明の実施に適したコンピュータ環境100の概略を図1を参照して説明する。一例としての動作環境100は、従来のパーソナルコンピュータシステム120を含む。このパーソナルコンピュータシステムは、プロセッサ121と、システムメモリ122と、システムバス123とを含む。システムバス123は、システムメモリ122をプロセッサ121に接続する。システムメモリ122は、リードオンリメモリ(ROM)124とランダムアクセスメモリ(RAM)125とを含む。ROM124は基本入出力システム126(BIOS)を含む。このBIOSは基本ルーチンを含む。この基本ルーチンは、起動時等にパーソナルコンピュータ120内の各要素間の情報転送を支援する。

【0037】パーソナルコンピュータシステム120は、ハードディスクドライブ127と、取出し可能磁気ディスク129を読み書きするための磁気ディスクドライブ128と、CD-ROM等の光媒体である取出し可能光ディスク131を読み書きするための光ディスクドライブ130とをさらに含む。ハードディスクドライブ127と磁気ディスクドライブ128と光ディスクドライブ130とは、各々ハードディスクドライブインタフェース132と磁気ディスクドライブインタフェース133と光ドライブインタフェース134とを介して、システムバス123に接続する。これらドライブおよび関連するコンピュータ読取り可能媒体は、不揮発性記憶手段をパーソナルコンピュータシステム120に提供する。ここに示したコンピュータ読取り可能媒体は、ハードディスク、取出し可能磁気ディスク、およびCD-R

OMディスクだが、当業者には明らかなように、他のコンピュータ読取り可能媒体を、一例として示した前記動作環境に用いても良い。例えば磁気カセット、フラッシュメモリカード、デジタルビデオディスク、ベルヌーイカートリッジ等を用いても良い。

【0038】ユーザは、キーボード140やマウス142等のポインティングデバイス等の入力装置を介して、パーソナルコンピュータシステム120にコマンドや情報を入力する。マイクロフォン161は、コンピュータシステム120に会話等の音声入力を行うために使用する。ユーザは、コンピュータシステムに図や手書きのグラフィック情報を入力する場合、スタイラスを用いて筆記タブレット162にグラフィック情報を描く。コンピュータシステム120は、他の入力装置(図示せず)としてジョイスティック、ゲームパッド、衛星アンテナ、スキャナ等を含んでも良い。マイクロフォン161は、システムバスに接続したオーディオアダプタ160を介してプロセッサ121に接続できる。他の入力装置は、システムバスに接続したシリアルポートインタフェース146を介してプロセッサ121に接続する場合が多い。しかしながらゲームポート、ユニバーサルシリアルバス(USB)等のインタフェースによって接続することもできる。

【0039】モニタ147等の表示装置は、ビデオアダプタ148等のインタフェースを介してシステムバス123に接続する。モニタに加え、パーソナルコンピュータシステムは、スピーカやプリンタ等の周辺出力装置(図示せず)を一般に含む。

【0040】パーソナルコンピュータシステム120は、ネットワーク環境で動作させ、1つ以上の遠隔コンピュータ149への論理接続を使用することもできる。遠隔コンピュータ149は、サーバ、ルータ、ピア装置、共通ネットワークノード等であり、一般に前記パーソナルコンピュータシステム120に関連して説明した要素の多くあるいは全てを含む。ただし図1は記憶装置150のみを示す。図1に示す論理接続は、ローカルエリアネットワーク(LAN)151と、ワイドエリアネットワーク(WAN)152とを含む。このようなネットワーク環境は、事務所、企業コンピュータネットワーク、イントラネット、およびインターネットにおいて普及している。

【0041】LAN環境で使用する場合、パーソナルコンピュータシステム120は、ネットワークインタフェース153を介してローカルネットワーク151に接続する。WAN環境で使用する場合、パーソナルコンピュータシステム120は、一般にモデム154等の手段を含み、インターネット等のワイドエリアネットワーク152を介して通信を確立する。モデム154は内蔵または外付けであり、シリアルポートインタフェース146を介してシステムバス123に接続する。ネットワーク

環境の場合、パーソナルコンピュータシステム120のプログラムモジュールは、遠隔記憶装置150に格納しても良い。図示のネットワーク接続は例であり、コンピュータシステム間の通信リンクは他の手段で確立しても良い。また本発明は、パーソナルコンピュータシステム以外に、ホストまたはサーバコンピュータシステムでも実現可能であり、CD-ROM以外の手段、例えばネットワーク接続インタフェース153によってホストコンピュータシステムにデータを送ることができる。

【0042】多くのプログラムモジュールをコンピュータシステム120の駆動装置やRAM125に格納できる。プログラムモジュールは、コンピュータシステム120の機能を制御し、ユーザ、入出力装置、あるいは他のコンピュータと対話する。プログラムモジュールは、ルーチン、オペレーティングシステム135、アプリケーションプログラムモジュール138、データ構造、ブラウザ、他のソフトウェア部品、あるいはファームウェア部品を含む。本発明は、テキスト音声化(TTS)モジュール137、入力インタフェースプログラムモジュール139等の1つ以上のプログラムモジュールにおいて好適に実施できる。これらモジュールの各々は、発明の詳細な説明に記載した方法に基づいている。

【0043】アプリケーションプログラムモジュール138は、本発明に関連して使用する様々なアプリケーションからなる。図2はそのうちのいくつかを示す。これらプログラムモジュールの目的および相互作用は、図2を参照して詳細に説明する。これらプログラムモジュールは、ワードプロセッサプログラム210(ワシントン州レドモンドのマイクロソフト社の製品であるワード等)、テキスト音声化モジュール137、第1音声認識プログラムモジュール240、および入力システム(IME)250を含む。

【0044】以下に説明する様々な手順を実行するための特定のプログラミング言語は説明しない。理由は、以下の説明および添付図面における動作、処理段階、および手順は、十分な開示を行っているため、当業者であればそれらに基づいて本発明の実施例を実施可能だからである。さらに実施例の実行に使用可能なコンピュータおよびオペレーティングシステムは多数あり、これら多くの異なるシステムに適用可能な詳細コンピュータプログラムを提示するのは困難だからである。特定コンピュータのユーザは、そのユーザの必要性および目的に最も適した言語およびツールを知っているであろう。

【0045】当業者には明らかなように、本発明は他のコンピュータシステム構成でも実施可能である。例えば携帯装置、マルチプロセッサシステム、マイクロプロセッサベースの消費者用電子機器、プログラム可能消費者用電子機器、マイクロコンピュータ、メインフレームコンピュータ等でも実施可能である。本発明は分散コンピュータ環境でも実施可能である。分散コンピュータ環境

では、通信ネットワークでリンクした各遠隔処理装置によってタスクを実行し、プログラムモジュールを局所記憶装置および遠隔記憶装置の両方に置くことができる。

【0046】(マルチソース入力再生ユーティリティの構成)図2は、マルチソース入力再生ユーティリティ200を示すブロック図である。マルチソース入力再生ユーティリティ200の目的は、複数のソースからユーザ入力を受け取り、その入力を処理してテキスト化し、そのテキストを音声として再生することである。入力ソースは、例えば手書き、音声、タイピングである。

【0047】マルチソース入力再生ユーティリティ200は、いくつかのプログラムモジュールを備える。これらプログラムモジュールは、様々なソースからワードプロセッサへのテキスト入力を可能にする。ユーザが入力したデータをテキストに変換し、ディスプレイ画面に表示する。ユーザは、音声データとして聞きたい部分のテキストを選択する。するとマルチソース入力再生ユーティリティ200は、適切なプログラムモジュールを介してその要求を処理し、音声を再生する。

【0048】マルチソース入力再生ユーティリティ200の代表的な入力ソースは、手書き認識プログラムモジュール230、第1音声認識プログラムモジュール240、入力システム(IME)250、第2音声認識プログラムモジュール240、および直接入力機器としてのキーボード140、マウス142等である。各入力プログラムモジュールを順次説明する。

【0049】手書き認識プログラムモジュール230は、ユーザから手書き入力280を受け取る。ユーザは、スタイラスやマウスで筆記タブレット162に手書きしたり、タッチスクリーンによって手書き入力280を生成する。手書き入力280は、オペレーティングシステム135の筆記タブレットドライバモジュールにより、手書き認識プログラムモジュール230に転送する。

【0050】第1音声認識プログラムモジュール240は、ユーザからの音声入力290を、マイクロフォン161を介して受け取る。このマイクロフォンは、オペレーティングシステム135のマイクロフォンドライバモジュールによって駆動する。音声は、同一発音で異なる意味や綴りを有する語がたくさんあるため、解釈が難しい場合が多い。第1音声認識プログラムモジュール240は、音声入力290を音素に分解し、各音素に基づいたテキスト出力を生成する。さらに第1音声認識プログラムモジュール240は、音声入力290を波形に変換し、それを保存音声データ270として長期記憶媒体に保存する。この保存音声データ270は、後から音声再生用としてワードプロセッサ210からアクセスする。

【0051】入力システム(IME)250は、音声入力を特定外国語の表意テキストに変換する。IME250への入力、例えばキーボード140やマウス142

を介してコンピュータに入力されるタイプしたテキストであり、あるいはマイクロフォン161を介して入力される音声データである。

【0052】IME250は、アジア言語の表意文字を生成する場合、特に有用である。これら言語は、キーボードのキーよりもはるかに多い表意文字を有するため、IME250が無ければ、コンピュータに特定の表意文字を入力するのは難しい。例えば中国語IME250の場合、ユーザは希望の中国語文字に対して英語文字音声スペルをタイプする。多くの中国語文字は発音が類似しているため、タイプした音声スペルは、多くの異なる漢字の1つを代表することがある。するとIME250は、内部的に生成した候補リストから最も可能性の高い候補を選択する。あるいはタイプした音声スペルに基づいて最適候補をかわるがわるユーザに提供し、ユーザは正しいものを選択できる。この候補リストはユーザが提供した音声スペルから、あるいはユーザの音声入力290に基づく音素セットから生成する。

【0053】IME250の機能を示すため、音声をコンピュータに入力しそれを使って例えば日本語テキストを生成してみる。その音声は第2音声認識プログラムモジュール240へ転送する。この第2音声認識プログラムモジュール240は、第1音声認識プログラムモジュール240とはほぼ同様に動作するが、独立したユニットとして図示した。その理由は、第2音声認識プログラムモジュール240は、異なる音声解釈エンジンと出力タイプを持つことができるからである。すなわち第2音声認識プログラムモジュール240は、第1音声認識プログラムモジュール240とは異なる言語を解釈できる。

【0054】第2音声認識プログラムモジュール240は、英語音声から1つ以上の代替英語テキストを生成する。この代替英語テキストは、IME250への入力となり、IME250はその英語テキストを日本語文字に変換する。IME250への各代替入力は、各別の候補リストを作るが、これら候補リスト間には重複があり得る。

【0055】IME250と第2音声認識プログラムモジュール240とは、直接にインタフェースしない。例えば第2音声認識プログラムモジュール240からIME250への入力は、インタフェースプログラムモジュールを介して送る。このインタフェースプログラムモジュールは、例えば入力インタフェース139である。この入力インタフェースは、各入力ソースに直接接続する。

【0056】英語と日本語の例を挙げたが、会話認識プログラムモジュール240とワードプロセッサ210は、複数言語の入力を受け取り、表示できる。例えばワードプロセッサ210は、ナスティリクのように、左から右以外の方向に読まれる言語テキストを表示する構成も可能である。またワードプロセッサ210は、ヘブラ

イ語やアラビア語のように、各文字の形が周囲の文字に応じて変化するテキストも表示できる。

【0057】ワードプロセッサ210は、ユーザからの入力を受け取り、入力インタフェース139を介して出力を表示する。この出力は、ディスプレイ画面上のテキスト、またはスピーカを介しての音声データである。ワードプロセッサ210は、マルチソース入力再生ユーティリティ200のユーザと、コンピュータプログラムモジュールとの間のインタフェースを提供する。

【0058】ワードプロセッサ210は、入力インタフェース139を介して、手書き入力280と音声入力290とに関して最適な代替テキストを受け取る。ある入力、例えばキーボード140でのタイピングによって生成されたテキストは、入力インタフェース139による代替テキスト生成を必要とせず、受け取ったままの形で転送される。ワードプロセッサ210は、全ソースデータを組合せ、マルチソーステキスト列を生成し、それをユーザに提供する。ワードプロセッサ210は、テキスト中の各ワードのソースをユーザに示すことはしないものの、ワードプロセッサは各テキスト要素のソースを記録する。別の実施例は、各語のソースを様々な方法で示すことができる。例えば各語を、その語の入力ソースに応じて異なる色やフォントで表示することができる。

【0059】ユーザは、ワードプロセッサ210を用いてテキストの一部を選択し、その選択部分の音声再生を要求できる。ワードプロセッサは、選択テキストに関連付けられた保存音声データ270を検索し、その音声データをスピーカ等の機器を介して再生する。選択テキストに関連付けられた保存音声データ270が無ければ、ワードプロセッサ210は、入力インタフェース139を介してテキスト音声化モジュール137に要求を送り、選択テキストに対応するTTS登録語を検索させる。ワードプロセッサはその後、第1音声認識プログラムモジュール240と入力インタフェース139を介して対応するTTS登録語を受け取り、それをスピーカを介して再生する。ワードプロセッサ210は、選択テキスト内の各語を解析し、保存音声データ270またはTTS登録語220を検索して再生する。マルチソース入力再生ユーティリティ200のユーザは、保存音声データとTTS登録語の組合せによる連続的な流れによる選択テキストの音読を聞く。選択テキストの音声再生の詳細は図5を参照して詳細に後述する。

【0060】(マルチソース入力再生ユーティリティ用ユーザインタフェース) 図3は、本発明の実施例に基づき口述文書サンプル300をディスプレイ画面147に表示したところを示す。図2の音声入力290は、第1音声認識プログラムモジュール240によって文字化され、ワードプロセッサ210によって口述文書サンプル300として表示される。本例の口述文書は、詩の一節である。第1音声認識プログラムモジュール240は、

口述の一部を誤認している。特に点線の四角で示す「白いフланネルのズボン (white flannel trousers: ホワイト フランネル トラウザー)」が「そりゃ日に焼けるから、ズボン (why tan I trouser: ホワイ タン アイ トラウザー)」となっており、第1音声認識プログラムモジュール240は誤ったテキストに変換している。

【0061】図4は、文書300を編集したものであり、部分310が音声再生用に選択されている。図4において、誤認部は編集により適切な語が挿入されている。すなわち「そりゃ日に焼けるから、ズボン (why tan I trouser)」を「白いフланネルのズボン (white flannel trousers)」に置き換えている。また文書部分310は、音声再生用に選択されている。文書300のうち、マルチソース入力再生ユーティリティ200によって音声化する部分をユーザにフィードバックするため、再生するテキスト部分310をハイライト表示する。再生用に選択したテキスト部分は、文書300の編集部分320を含んでいることに注意したい。その編集がキーボードのタイピングでなされていれば、文書300はマルチソース文書である。テキスト部分310はマルチソーステキスト部分である。「白いフланネルのズボン (white flannel trousers)」は関連付けられた音声データを持たない。なぜならキーボードを使って入力したからである。ユーティリティ200が再生を開始すると、選択テキスト部310の「白いフланネルのズボン (white flannel trouser)」は、テキスト音声化モジュール137が提供するテキスト音声化登録語220を使って音声再生される。

【0062】ユーザフィードバックをさらに強化するため、マルチソース入力再生ユーティリティ200は、音声再生する語の正確な指示を行うことができる。例えば音声データまたはテキスト音声化登録語の再生に従って、対応する語を点滅させたりその語の色を変えたりする。

【0063】(マルチソース入力再生ユーティリティの動作) マルチソース入力再生ユーティリティ200の各部動作を図5に基づき説明する。説明は図2の機能ブロック図を適宜参照しながら行う。

【0064】図5は、文書の選択部分に対する音声再生処理の詳細を示すフローチャートである。まずステップ400において、ユーザはマルチソース入力再生ユーティリティ200によって音声再生する文書のテキスト部分を選択する。再生用テキストを選択すると、ワードプロセッサ210はステップ405において、テキスト部分300の第1語の境界を決定する。一般に語の境界は、スペース、アスタリスク、カンマ、ピリオド等の非英数字である。しかしながら言語によっては区切り文字がない。その場合、語の区切りを決定するための言語理

解形式を用いねばならない。各語の区切りが無い言語の例は、日本語、中国語、タイ語である。以下の説明は、区切られた語を「現在の語」と呼ぶ。

【0065】ワードプロセッサ210が現在の語の大きさと境界とを決定すると、マルチソース入力再生ユーティリティ200は、ステップ410において、その語が、先の口述において保存した保存音声データ270に関連付けられているかを決定する。

【0066】ステップ410においてその語が保存音声データ270に関連付けられていれば、マルチソース入力再生ユーティリティ200は、その音声データをステップ415において取り出す。ワードプロセッサ210は、その保存音声データを入力インタフェース139に要求し、該入力インタフェースはそのデータを磁気ディスクドライブ128等の媒体から取り出す。入力インタフェース139はその音声データをワードプロセッサ210に渡す。

【0067】ステップ415の後、マルチソース入力再生ユーティリティ200はステップ420において、前記音声データを再生する。一般にワードプロセッサ210は、独立したプログラムモジュールあるいはオペレーティングシステムの一部を用いて音声データを再生する。他の実施例において、ワードプロセッサは再生プログラムモジュールを内蔵できる。

【0068】マルチソース入力再生ユーティリティ200は次に、ステップ405～420で処理した語が、ステップ400においてユーザが選択したテキストの最後の語であるかを決定する。選択範囲に語が残っていなければ、ステップ460において終了する。語が残っていれば、マルチソース入力再生ユーティリティ200はステップ405に戻り、次の語の境界を決定する。

【0069】ユーティリティ200がステップ410において、現在の語に関連付けられた音声入力はないと決定した場合を説明する。例えば文書編集処理において、ユーザがタイプした語をユーティリティ200が処理する場合である。この場合、ユーティリティ200は、現在の語に対応するTTS登録語220があるか否かをチェックする。すなわちワードプロセッサ210は、第1音声認識プログラムモジュール240を介してTTSモジュール137に対し、現在の語に対応するTTS登録語があるか否かを問い合わせる。

【0070】対応するTTS登録語220があれば、TTSモジュール137はそれを取り出し、入力インタフェース139を介してワードプロセッサ210に渡す。あるいはTTSモジュール137は、取り出すTTS登録語を決定するに当たり、現在のTTS登録語の両側の音素に対応する周囲のTTS登録語を調べ、周囲のTTS登録語に音素的に最も一致するものを選択する。TTSモジュール137は、この処理をステップ430において行う。

【0071】ワードプロセッサがTTS登録語を受け取ると、マルチソース入力再生ユーティリティ200は、ステップ420においてそれを再生する。その詳細は保存音声データ270の再生で説明したとおりである。あるいはステップ420における再生前に、TTS登録語のピッチ、トーン、速度等の可聴特性をユーティリティ200によって処理し、TTS登録語の音声を保存音声データに近似させ、2タイプの音声再生間の不調和を最小にしても良い。TTS登録語220を再生した後、ステップ455を前記通りに実行する。

【0072】現在の語に対応するTTS登録語が無ければ、ワードプロセッサ210はステップ435において、その語を音素に分解する。ステップ440においてワードプロセッサは、第1音素を入力インタフェース139を介して第1音声認識プログラムモジュール240に転送する。ステップ440の一部として、音声認識プログラムモジュールは、テキスト音声化モジュール137に対し、第1音素に対応するTTS登録語220を要求する。特に第1音声認識プログラムモジュール240は、ワードプロセッサ210から受け取った音素を分析し、TTSモジュールに対し、その音素の最適な異音からなるTTS登録語220を要求する。TTS登録語を取り出しそれをワードプロセッサ210に転送してステップ440を終了する。

【0073】（結論）マルチソース入力再生ユーティリティ200は、ユーザが手動で再生速度を変えられる等の追加機能を含んでも良い。ユーティリティ200は、前記した以外の入力ソースからの入力を受け付けても良い。例えば別の入力装置や異なるプログラムからの制御信号を受け付けることができる。以上説明した本発明の実施例は、それに基づいて多くの変更や追加を可能にす

る。すなわち前記説明は本発明の特定の実施例のみに関しており、本発明には、請求の範囲に記載の本発明範囲を逸脱することなく、多くの変更が可能である。

【0074】

【発明の効果】以上述べたように、本発明によれば、文書の選択部分に対応する音声を直感的な形式で確実に再生する方法およびシステムが提供できる。また、音声データが欠落した部分を有する文書の音声再生において、その欠落部分を穴埋めする方法およびシステムが提供できる。

【0075】すなわち、本発明によれば、音声をテキストに関連付ける方法、コンピュータで文書を生成し読み上げる方法、コンピュータでテキスト文書の音声再生を行う方法、及び、文書内のテキストを編集し評価する方法を提供できる。

【0076】本発明によれば、音声をテキストに関連付けるコンピュータ、文書を生成し読み上げるコンピュータ、及び、テキスト文書の音声再生を行うコンピュータを提供できる。

【図面の簡単な説明】

【図1】本発明の実施例を動作させる環境の一例としてのパーソナルコンピュータを示すブロック図である。

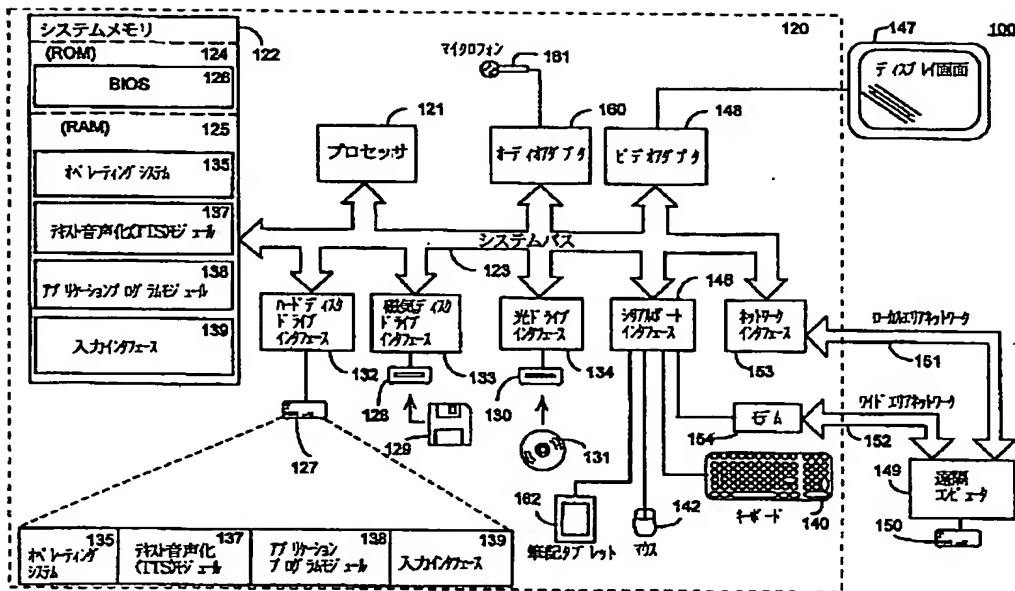
【図2】マルチソース入力再生システムのソフトウェア構成を示すブロック図である。

【図3】マルチソース入力再生システムを使用して入力したテキスト文書の一部を示すディスプレイ画面である。

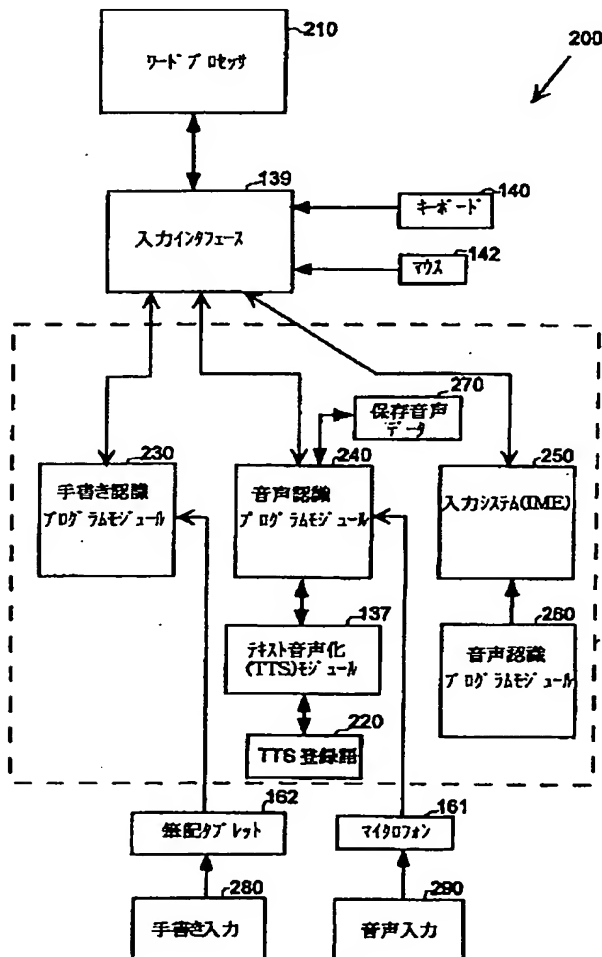
【図4】編集済みテキスト文書の中で音声再生部分を選択したところを示すディスプレイ画面である。

【図5】マルチソース文書の一部を選択し音声化する方法を示す論理フローチャートである。

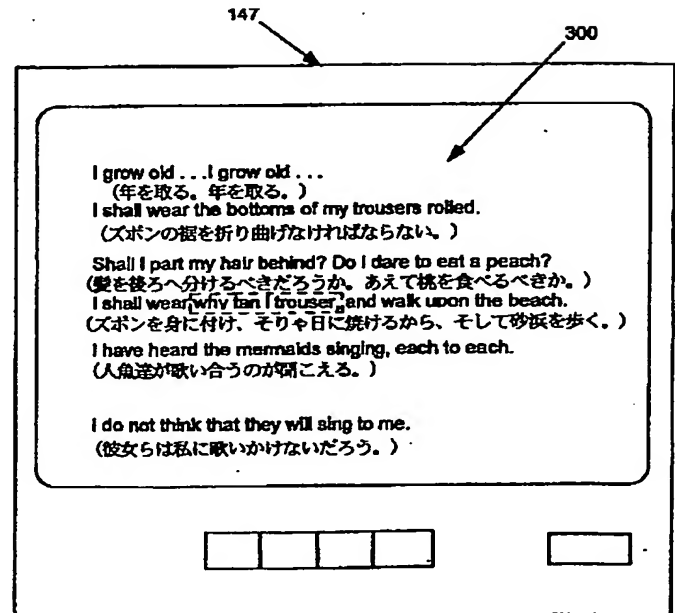
【図1】



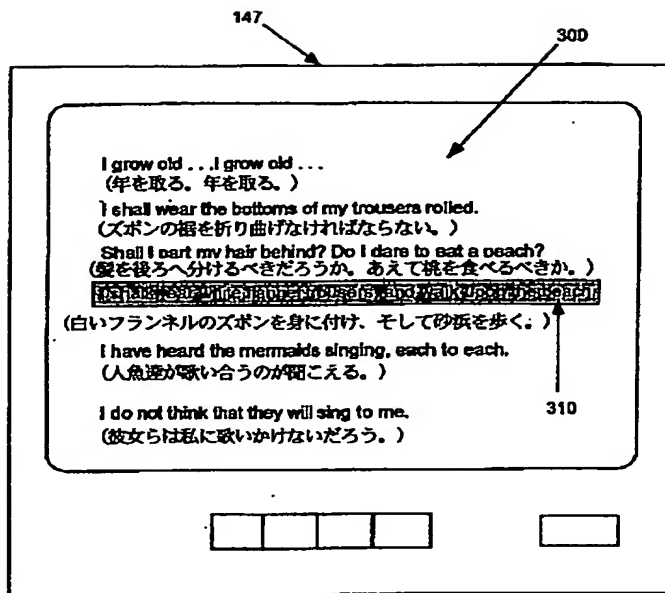
【図2】



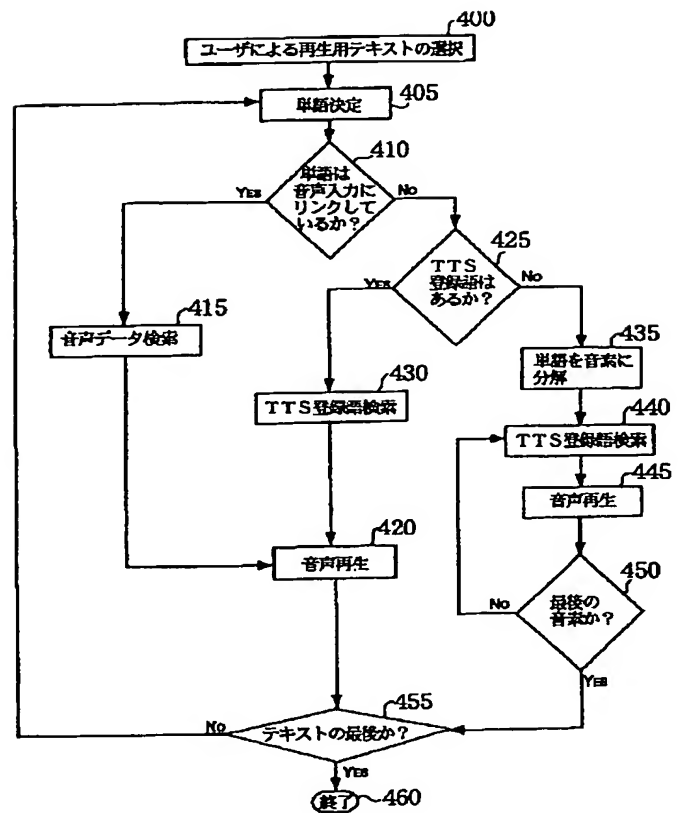
【図3】



【図4】



【図5】



フロントページの続き

(51)Int.Cl.⁷
G 1 0 L 15/00

識別記号

F I
G 1 0 L 3/00

テーマコード(参考)

5 5 1 B

(72)発明者 エリック ラッカー
アメリカ合衆国 98105 ワシントン州
シアトル フォーティエイス アヴェニュー
ー 4316

(72)発明者 ポール キョン ホワン キム
アメリカ合衆国 98107 ワシントン州
シアトル ナンバー205 エヌダブリュー
サーティナイン ストリート 201

(54)【発明の名称】 音声をテキストに関連付ける方法、音声をテキストに関連付けるコンピュータ、コンピュータで文書を生成し読み上げる方法、文書を生成し読み上げるコンピュータ、コンピュータでテキスト文書の音声再生を行う方法、テキスト文書の音声再生を行うコンピュータ、及び、文書内のテキストを編集し評価する方法